BIOLOGICALLY MOTIVATED SPIRAL ARCHITECTURE FOR FAST VIDEO PROCESSING

BIOLOGICALLY MOTIVATED SPIRAL ARCHITECTURE FOR FAST VIDEO PROCESSING

Min Jing[1*], *Sonya Coleman*[1], *Bryan Scotney*[2]

[1]Intelligent Systems Research Centre
[2]School of Computing and Information Engineering
University of Ulster
United Kingdom

Martin McGinnity

School of Science and Technology
Nottingham Trent University
United Kingdom

ABSTRACT

Fast image processing is a key element in achieving real-time image and video analysis. The spiral addressing scheme [10] has been an efficient tool for hexagonal image processing (HIP), whereby the image pixel indices are stored in a one-dimensional vector that enables fast processing. Unlike HIP, which requires a complex resampling scheme, we present a novel "squiral" (square spiral) image processing (SIP) framework that provides a spiral addressing scheme for direct application to standard square pixel-based images. A SIP-based non-overlapping convolution technique is developed by simulating the eye tremor phenomenon of the human visual system to accelerate computation in feature extraction. Furthermore, we deploy the proposed simulated eye tremor technique on a sequence of video frames. The preliminary results based on two action video clips demonstrate the potential of the SIP-based eye tremor model to facilitate fast video processing.

Index Terms— spiral image processing, spiral addressing, eye tremor, non-overlapping convolution, video processing

1. INTRODUCTION

Real-time data processing is a challenging task, particularly when handling large-scale image and video data from social media. Recently fast image processing based on a hexagonal image processing (HIP) framework [5, 7] has attracted attention [1,2,4,9]. By incorporating a spiral architecture [10], the HIP image pixel indices can be stored in a one-dimensional vector that enables fast processing. However, the computational advantages of HIP are undermined significantly by the additional time and effort required for conversion of standard image data to a HIP environment, as existing hardware for image capture and display are based predominantly on traditional rectangular pixels.

In this work, we introduce a novel "squiral" (square spiral) image processing (SIP) framework that develops a novel spiral addressing scheme for standard square pixel-based images. Hence, unlike HIP, conversion of (standard two-dimensional) pixel indeces to the SIP addressing scheme can be achieved easily using an existing lattice with a Cartesian coordinate system. Further, the approach can be used to implement efficiently existing image processing operators designed for standard rectangular pixel-based images; there is also no need to design special image processing operators, as was the case for HIP [2].

Standard feature extraction is usually executed via convolution, where typically a gradient-based operator is applied to a pixel and its neighbours. The traditional approach to feature extraction based on overlapping convolution operators does not closely represent the human visual system. Furthermore, the human visual system does not process single static images, but instead a series of temporal images that are slightly off-set due to involuntary eye movements. The use of eye tremor, rhythmic oscillations of the eye, for image processing was first exploited in [8] to mimic the human visual system and potentially reduce the heavy computational burden of standard convolution. To more closely mimic the human visual system, a biologically inspired framework was proposed [9] by modelling eye tremor for HIP using the one-dimensional addressing scheme of the spiral architecture. Inspired by [9], we developed a SIP-based non-overlapping convolution technique by simulating the eye tremor phenomenon to facilitate fast computation. For application, we deploy the proposed approach to a sequence of video frames to investigate its potential for fast video processing.

We first explain the spiral addressing scheme for square pixel-based images and conversion to SIP from the standard Cartesian 2D addressing scheme. We then demonstrate the development of SIP-based non-overlapping convolution. In experiments the SIP-based eye tremor model is applied to video frame sequences representing both very small (almost imperceptible) and perceptible degrees of movement between successive frames. The preliminary results demonstrate the potential of the proposed method for efficient video processing.

2. SPIRAL IMAGE PROCESSING

2.1. Spiral Addressing Scheme

The proposed spiral scheme for square pixel-based images is inspired by the spiral addressing for HIP [10]. Similar to HIP,

*This work is supported by FP7 project SLANDAIL.

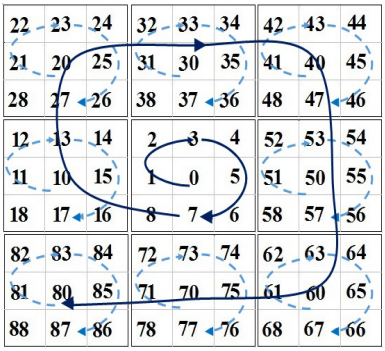


Fig. 1. The spiral addressing scheme for layer-2 SIP.

0	1	2	3	4	5	6	7	8	10	11	...	18	20	...	28	30	...

Fig. 2. One-dimensional address values for a SIP image.

the SIP image originates at the centre of a square image and spirals out using one-dimensional indexing. As illustrated in Fig. 1, a layer-1 SIP cluster consists of nine pixels (ie, nine layer-0 SIP clusters); higher order layers are generated recursively, i.e., a layer-2 SIP cluster consists of nine layer-1 SIP clusters, comprising 81 pixels in total. The SIP image is stored in a one-dimensional vector based on the spiral addressing scheme as illustrated (in part) in Fig. 2.

2.2. Conversion to SIP from a Square Image

Unlike HIP conversion which requires pre-processing using a resampling scheme to match the location and value of points in an original rectangular pixel-based image to location and value in a hexagonal lattice [7], SIP conversion can use directly the original lattice of a square image. For a rectangular image with size $M \times N$, the number of SIP layers λ can be found by $\lambda = (logM + logN)/log9$; then the length of the (one-dimensional) SIP image is 9^{λ}. We can adapt the spiral addressing scheme for HIP [10] to the SIP case, and thus a SIP address can be represented as:

$$a_n a_{n-1} ... a_1 = \sum_{i=1}^{n} a_i \times 10^{i-1} \tag{1}$$

where $0 \leq a_i < 9$. $\sum$ denotes Spiral Addition and $\times$ indicates Spiral Multiplication [7]. For example, the point at SIP address 867 can be located by accumulating the shifts corresponding to the locations of the addresses of 800, 60 and 7. To locate a SIP address corresponding to Cartesian coordinates (x,y) in a standard square image, we define the centre of the image as $L(0) = (0, 0)$. Based on the SIP addressing scheme, we have: L(1)=(-1,0), L(2)=(-1,1), L(3)=(0,1), L(4)=(1,1), L(5)=(1,0), L(6)=(1,-1), L(7)=(0,-1) and L(8)=(-1,-1). To locate the points in a higher SIP layer, we calculate of the shift required from the centre point to the target point by

$$L(a_i \times 10^{i-1}) = 3^{i-1} \times L(a_i) \tag{2}$$

For example, the point L(87) can be located by $L(87) = L(80)+L(7) = 3\times L(8)+L(7) = 3\times(-1,-1)+(0,-1) = (-3,-4)$. Similarly, a point at L(4536) can be located by $L(4536) = L(4000)+L(500)+L(30)+L(6) = 3^3 \times L(4)+3^2 \times L(5) + 3 \times L(3) + L(6) = (37, 29)$. Hence the point L(4536) can be found by shifting the start point from (0,0) to (37,29). After conversion from a square image, the SIP image is stored as a vector for further processing.

3. SIP CONVOLUTION VIA EYE TREMOR

3.1. Simulation of Eye Tremor

When using a standard 2D addressing scheme on an image, the addresses of a pixel's neighbours can be determined easily. However, determining a pixel's neighbours in a one-dimensional addressing scheme is not straightforward and requires significant computation. Inspired by [9], we alleviate this computational burden by developing an eye tremor based framework for SIP. The pixel offsets in a layer-1 eye tremor are illustrated in Fig.3 which includes nine offset images. Consider I_0 as a "base" SIP image; eight additional images $I_j, j = 1, 2, ..., 8$ is obtained by shifting I_0 by one pixel in the image plane along the spiral addressing scheme. The "centre" of each image I_j is located at a pixel within the layer-1 neighbourhood centred at image I_0. Each image is stored as a vector after being converted from the 2D image structure.

I_2	I_3	I_4
I_1	I_0	I_5
I_8	I_7	I_6

Fig. 3. Pixel offsets for a layer-1 eye tremor using 9 images.

3.2. SIP-based Convolution

Feature detection operators are often based on first derivative approximations. Unlike HIP-based feature extraction, which requires x- and y-components of a hexagonal operator to be designed accordingly [2], the SIP-based approach supports direct convolution of standard image processing operators after converting them to a SIP vector. For this study, we use Sobel (a layer-1 operator) as the example for edge detection. For a given image I_0, convolution of a Sobel operator (denoted as H_1) across the entire image plane is achieved by applying the operator sparsely to each of the nine images $I_j, j = 0, ..., 8$ and then combining the resultant outputs. Based on the eye tremor framework, for each image I_j, we apply the operator H_1 only when centred at those pixels with spiral address 0(mod 9), hence achieving non-overlapping convolution. The convolution of the image I_j with an operator H_λ can be defined as,

$$G_\lambda^j(s_0) = \sum_{s \in N_\lambda(s_0)} H_\lambda(s) \times I_j(s), \tag{3}$$

where $\forall s_0 \in \{s | s = 0(mod 9))\}$ and $N_\lambda(s_0)$ denotes the λ-neighbourhood centred on the pixel with spiral address s_0 in

image I_j. For layer-1 operator H_1, the matrix implementation of convolution with I_0 in Eq. (3) can be written as:

$$\begin{pmatrix} G_1^0(0) \\ G_1^0(10) \\ \vdots \\ G_1^0(k) \end{pmatrix} = \begin{pmatrix} I_0(0) & I_0(1) & ... & I_0(8) \\ I_0(10) & I_0(11) & ... & I_0(18) \\ \vdots & \vdots & \ddots & \vdots \\ I_0(k) & I_0(k+1) & ... & I_0(k+8) \end{pmatrix} \begin{pmatrix} H_1(0) \\ H_1(1) \\ \vdots \\ H_1(8) \end{pmatrix} \tag{4}$$

where $k = 0, 10, 20, 30, ...$ We can apply the same process to the remaining eight images ($I_1, ..., I_8$); each is an image created by shifting the origin by one pixel from I_0. The overall outcome can be obtained by assembling the values of G_1^j into a vector with the following arrangement:

$$[G_1^0(0)G_1^1(0)...G_1^8(0)G_1^0(10)G_1^1(10)...G_1^8(10)...G_1^0(k)...G_1^8(k)] \tag{5}$$

This assembly within the one-dimensional vector achieves the same outcome as standard 2D convolution. However, since all processing is executed in vector form, computation is significantly faster than standard convolution.

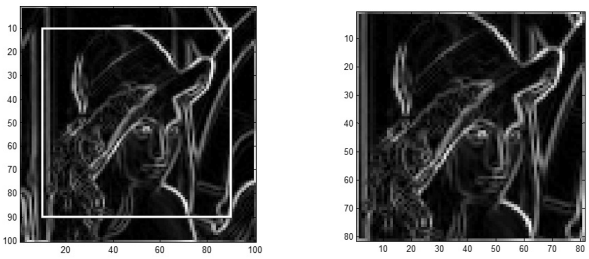
4. APPLICATION TO VIDEO PROCESSING

Motivated by the real-time processing capabilities of the human visual system, we consider how characteristics of that system can simulated to reduce computational effort when implementing low-level feature extraction. As a consequence of eye tremor, through rhythmic oscillations of the eye the human visual system processes series of temporal images that are slightly offset [8]. Therefore, we use a set of similarly offset images (video frames) that are each partially processed by non-overlapping filters as described in Section 3. We thus extend the proposed SIP approach to video processing, in which SIP is applied to each video frame. For layer-1 eye tremor, we consider nine consecutive video frames as nine eye tremor images as shown in Fig.3, with each frame having a different offset sequentially. For performance evaluation we compare the results from combined nine consecutive frames by SIP technique with the those obtained from the ninth frame by full (overlapping) convolution. This comparator represents the output from traditional convolution of the edge detection operator at every pixel, where full convolution would also have been applied to all previous frames. The difference is that in traditional convolution the operator is applied at every pixel in every frame, whereas in the SIP approach the operator is applied at only one-ninth of the pixels in each frame.

5. EXPERIMENT RESULTS

5.1. Single Image Edge Detection

The result of SIP-based edge detection for the lena image is shown in Fig. 4(b); Fig 4(a) shows the output from standard 2D convolution. In Fig 4(a) the SIP region is outlined, which corresponds to the whole image shown in Fig 4(b). As SIP and standard convolution differ only in the approaches used for pixel indexing and image storage, the results of convolution are the same for both methods. This can be seen in the



(a) Standard 2D convolution (b) SIP convolution

Fig. 4. Feature maps

feature maps in Fig 4. The size of the SIP image is slightly smaller because SIP is based on a square image that corresponds in size to the layer-4, and hence a border region beyond that may not be processed. This is not a significant restriction, as in most video applications the elements of interest are unlikely to be at the periphery of the image. Because the SIP image is stored as a vector, all of the computational processes can be executed more efficiently than in the standard 2D convolution approach as discussed in [6]. (We do not include the cost of SIP conversion and extra memory).

5.2. Video Frame Images

For application to video data we deploy the eye tremor concept to video frame images by considering each nine consecutive frames as eye tremor images, and we compare the edge feature output with that obtained by application of the standard full convolution to the ninth frame, as discussed in Section 4. To evaluate the performance under different degrees of motion within the video, we selected two short video clips from the Weizmann [3], including two actions: "jump" and "walk". To ensure that only small movements occur between consecutive frames, we repeated each frame to increase the frame rate from 25 frames/sec to 50 frames/sec (though, of course, this does not provide an accurate representation of the action). Each video frame has a dimension of 250×250, and which was then converted to a sequence of layer-5 SIP image (corresponded to a 243×243 image.) For each action we further selected two sets of nine consecutive frames as shown in Fig. 6: (a) the first set contained little visually perceptible movement, whilst (b) the second set contained clearly perceptible changes between frames.

The output from the first set for "jump" and "walk" are shown in Fig. 7 and Fig. 9, respectively. In each case the output from standard full convolution on the ninth frame are shown in (a) and those from application of SIP to frames 1-9 are shown in (b). For both "jump" and "walk" the edge feature in (a) and (b) are visually similar. For the second set of frames, results for "jump" and "walk" are given in Fig.8 and Fig.10. There are more obvious differences between the outputs in (a) and (b), with (b) retaining a "history" of the movement across the nine frames to which SIP has been applied.

The results shown in Fig. 7 and Fig. 9 suggest that when

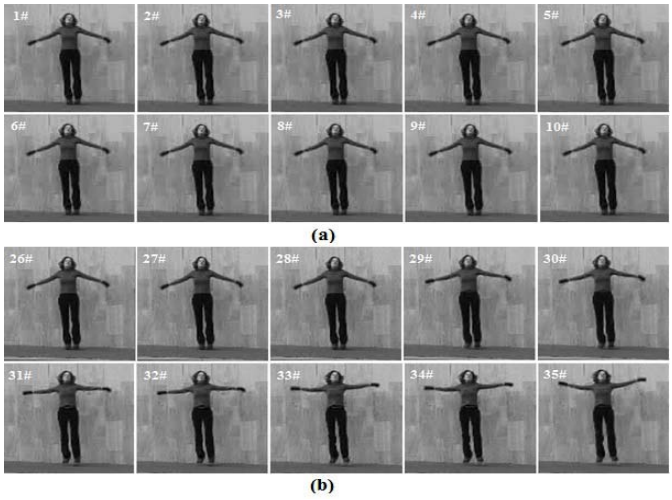


Fig. 5. Selected video frame sets for action "jump: (a) the 1st set (frames 1-9); (b) the 2nd set (frames 26-34).

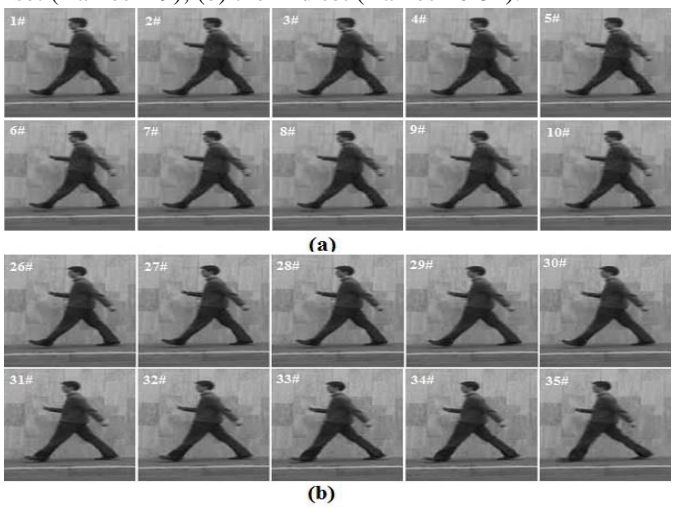


Fig. 6. Selected video frame sets for action "walk: (a) the 1st set (frames 1-9); (b) the 2nd set (frames 26-34).

movement between frames is small, the continuous, but partial processing of all nine video frames using SIP can provide edge feature output of a quality that is visually similar to that provided by full convolution on the 9th frame. When there are greater movement, as across frames 26-34, the edge feature output differs between full convolution on every frame and application of SIP (as in Fig. 8 and Fig. 10), with SIP yielding a more blurred output that represents motion across a sequence of nine frames. The current approach has been based on layer-1 eye tremor only, involving offset of just one pixel between successive video frames, and this is appropriate for application of small operators such as Sobel. For future extension of SIP we will consider extending the eye tremor model to higher levels that are appropriate to operators at larger-scale. With SIP then available over a range of layers, it is possible to implement adaptive multi-scale feature extraction on video data. The results presented here based on layer-1 SIP case are promising and suggest the potential of

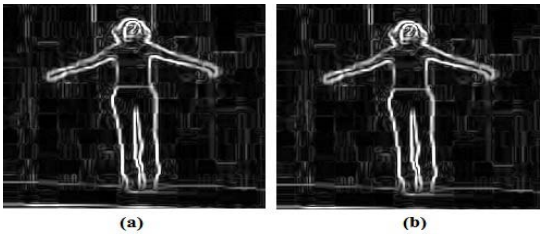


Fig. 7. Edge detection obtained from: (a)frame 9; (b) frame 1-9.

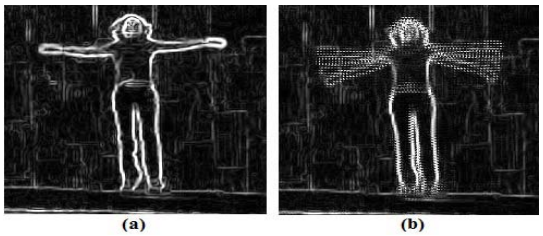


Fig. 8. Edge detection obtained from: (a) frame 34; (b) frame 26-34.

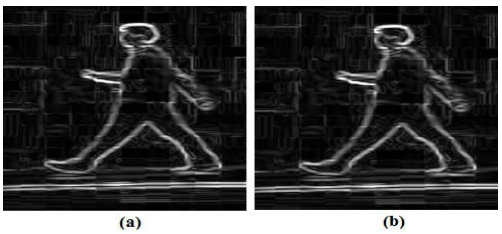


Fig. 9. Edge detection obtained from: (a)frame 9; (b)frame 1-9.

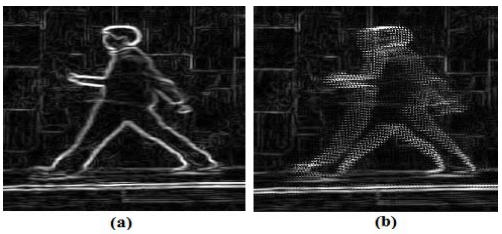


Fig. 10. Edge detection obtained from: (a) frame 34; (b) frame 26-34.

the proposed method for fast video processing.

6. CONCLUSION

We have introduced a novel spiral image processing framework for standard square pixel-based images. By incorporating a spiral architecture in conjunction with eye tremor, a non-overlapping convolution is developed to facilitate fast processing. We have deployed the concept of eye tremor to video data analysis by considering a sequence of video frames as a set of eye tremor images. The preliminary results demonstrate the potential of the proposed method for fast video processing. Further extension will use higher order SIP layers for implementation of adaptive multi-scale operators for feature extraction in real-time image and video content retrieval.

7. REFERENCES

[1] F. Asharindavida, N. Hundewale, S. Aljahdali,"Study on Hexagonal Grid in Image Processing", In Proc. ICIKM 2012.

[2] SA. Coleman, BW. Scotney and B. Gardiner,"A Biologically Inspired Approach for Fast Image Processing", In IAPR Proc. Machine Vision Applications, pp.129-132, 2013.

[3] L. Gorelick, M. Blank, E. Shechtman, M. Irani and Ronen Basri, "Actions as Space-Time Shapes", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.29, no.12, pp. 2247-2253, 2007.

[4] X. He, et al.,"An Approach to Edge Detection on a Virtual Hexagonal Structure", Digital Image Computing Techniques and Applications, pp. 340-345, 2007.

[5] I. Her, "Geometric transformations on the hexagonal grid", IEEE Transactions on Image Processing, 4(9), pp. 1213- 1222, 1995.

[6] M. Jing, BW. Scotney, SA. Coleman and TM. McGinnity, "Biologically Inspired Spiral Image Processing for Square Images", In Proc. IAPR MVA 2015.

[7] L. Middleton and J. Sivaswamy,"Hexagonal Image Processing; A Practical Approach", Springer 2005.

[8] A. Roka, et al. "Edge Detection Model Based on Involuntary Eye Movements of the Eye-Retina System", Acta Polytechnica Hungarica, 4(1), pp31-46, 2007.

[9] BW. Scotney, SA. Coleman and B. Gardiner, "Biologically Motivated Feature Extraction Using the Spiral Architecture", In Proc. IEEE ICIP, pp.221-224, 2011.

[10] P. Sheridan, T. Hintz and D. Alexander,"Pseudo-invariant Image Transformations on a Hexagonal Lattice", Image and Vision Computing, vol. 18, pp. 907-917, 2000.